ACS VENTURES
BRIDGING THEORY & PRACTICE

Evaluation of the Alignment of the Nebraska Alternate Assessment to the Nebraska Alternate Mathematics Standards

March 18, 2018

Submitted by:
Susan Davis-Becker, Ph.D.
410-845-7540
sdavisbecker@acsventures.com

Contents

Overview and Purpose

In 2015, the Nebraska Department of Education (NDE) adopted revised expectations for students in Mathematics in the Nebraska's College and Career Ready Standards for Mathematics. These standards are organized into four domains (Number, Algebra, Geometry, and Data) with subsumed grade level expectations (GLEs) and curricular indicators that detailed the specifics for each GLE by grade level. During the 2017-2018 school year, the NDE adopted a revised set of specifications for the Nebraska Student-Centered Assessment System (NSCAS) Alternate Assessment in Mathematics (NSCAS – AA M) which included extended indicators that were individually matched to the curricular indicators from the 2015 standards along with targets for the number of items that would be included on each test form aligned to each extended indicator (see Appendix A).

In January 2018, ACS Ventures, LLC (ACS) facilitated an alignment study on behalf of the NDE. The purpose of this study was to evaluate the alignment between the redeveloped NSCAS – AA M and the Nebraska Extended Indicators in Mathematics for grades 3-8 and 11. The purpose of this report is to document the approach taken to collecting alignment information, the analysis of the results, and the conclusions that can be drawn based on these results.

Alternate Alignment Approach

The alignment study for the NSCAS – AA M was designed to follow the *Links for Academic Learning* methodology (LAL method; Flowers, Wakeman, Browder, & Karvonen, 2009). The purpose of alignment in this context is to evaluate the connections among the knowledge, skills, and abilities (KSAs) in Nebraska's College and Career Ready Standards for Mathematics, the associated Extended Indicators, and assessed by the Alternate Assessments (AA items). The LAL method requires panelists to make judgments about two links within this system:

- Link 1: The connection between the Extended Indicators and the Nebraska's college and career ready standards for Mathematics
- Link 2: The connection between the content of the Alternate Assessment in Mathematics and the Nebraska Extended Indicators in Mathematics.

Within the LAL method, there are several critera by which the two links are evaluated.

- Criteria 1 – the content is acadmeic
- Criteria 2 – the content is referenced to the students' assigned grade levels
- Criteria 3 – the focus of achievement maintains fidelity with the content of the original grade-level standards (content centrality), and when possible, the specified performance (performance centrality)
- Criteria 4 – the content differes from grade level in range, balance, and DOK but matches high expectations for students with significant cognitive disabilities. [Link 1 only]
- Criteria 5 – there is some differentiation in content across grade levels

Each of these criteria is further delinated in Table 1 which provides the specific alignment questions for each of the links as well as the alignments ratings/evidence that is collected to answer each question and evlauate the overall criteria. The alignment study activities were desigend to collect the evidence outlined in the table as a series of judgments from submect matter experts (SMEs).

Table 1. Alignment Evaluation Quesitons and Evidence by Criteria and Link

Criteria & Link	Alignment Question	Alignment Ratings/Evidence
1. The content is academic		
1	Do the extended indicators measure academic content (KSAs)?	Number and percentage of extended indicators that are rated as academic or nonacademic/ foundational
2	Do the AA items measure the KSAs that are included in extended indicators?	Number and percentage of AA items that are rated as academic or nonacademic/foundational
2. The content is referenced to the assigned grade level		
1	Do the extended indicators reference KSAs that are included in the associated grade-level curricular indicators?	Number and percentage of extended indicators that are referenced to grade-level curricular indicators.
2	Do the AA items measure the KSAs that are included in the grade-level extended indicators?	Number and percentage of AA items that are referenced to grade-level extended indicators.
3. The focus of achievement maintains fidelity with the content of the original grade-level curriuclar indicators (content centrality), and when possible, the specified performance (performance centrality).		
1	Do the extended indicators maintain fidelity with the content and performance expectations of the original grade-level curricular indicator?	Number and percentage of extended indicators - rated as near, far, or no match to the content of the grade-level curricular indicators - rated as having all, some, or none of the same performance of the grade-level curricular indicators
2	Do the AA items maintain fidelity with the content and performance expectations of the extended indicators?	Number and percentage of AA items - rated as near, far, or no match to the content of the grade-level extended indicators - rated as having all, some, or none of the same performance of the grade-level extended indicators
4. The content differs from grade level in range, balance, and DOK but matches high expectations for students with significant cognitive disabilities.		
2	Categorical concurrence: Do the same categories of content appear in the extended indicators and the AA items? Depth of knowledge: Do the AA items target the same range of cognitive levels as the extended indicators?	Categorical concurrence: the distribution of how the AA were aligned to the Domains, GLEs, and extended indicators. Depth of Knowledge: the distribution of AA items across the DOK stages and correspondence to the target cognitive level/levels for the extended indicators.

Criteria & Link	Alignment Question	Alignment Ratings/Evidence
	Range of knowledge: Do the AA items target the same range of knowledge as the Domains and GLEs? Balance of representation: Are the AA items distributed across the GLEs within each Domain?	Range of knowledge: the percentage of GLEs, within each Domain, that are represented by AA items. Balance of representation: the degree to which one GLE is given more emphasis on the assessment than another within a Domain.
5. There is some differentiation in content across grade levels.		
1	How does the content specified in the extended indicators change across grade levels?	Number and percentage of extended indicators that specify: - broader classification of the target skill or knowledge - deeper mastery of the target skill or knowledge - a skill that for which the prerequisite was included at a lower grade standard - a new skill or knowledge that was not included in the lower grade levels - a skill or knowledge that is identical to one that was included in a lower grade level
2	Are the AA items age appropriate for each grade level?	Number and percentage of AA items that are identified as: - Adapted from age-specific content - Age neutral - Inappropriate for age group

Study Process

NDE recruited 15 SMEs to serve as panelists on each of three panels (5 SMEs per panel): grades 3 and 4, grades 5 and 6, and grades 7, 8, and 11. One member of each panel was selected to be the table leader and help facilitate consensus discussions and record judgments. The panelists were selected from 15 school districts across the state based on their experience and expertise. A summary – at the panel level – is shown in Table 2.

Table 2. Panelist Demographic Information

		3 - 4	5 - 6	7, 8, 11
Number of Panelists		5	5	5
Job Role				
	Special Education Teacher	1	1	3
	Director/Supervisor	2	3	2
	Professor	1		
	Instructional Coach	1		
	School Psychologist		1	

	3 - 4	5 - 6	7, 8, 11
Average Years of Experience	26	25	14
Highest Earned Degree			
Bachelors			2
Masters	5	3	2
Ph.D.		2	1

ACS prepared the materials for the study including training documents and rating forms (see Appendix A). NDE provided access to the Nebraska Extended Indicators in Mathematics and the Data Recognition Corporation (DRC) provided access to the NSCAS – AA M. The meeting process was conducted as follows. On the first day, Jeremy Heneger from the NDE welcome the panelists, reviewed the purpose of the alignment study, and answered questions from the panel about alternate assessment in Mathematics. Dr. Susan Davis-Becker from ACS then facilitated a training on the alignment process, the judgments that the panel would be asked to make, and how the results were to be used. The panelists then worked in their grade-level panels for the remainder of the study by reviewing their assigned materials and making the judgments outlined by the facilitator. When needed, staff from the NDE were available to answer questions. The meeting concluded on the second day after the judgmental process was complete and panelists had responded to an evaluation of the process.

To gather the evidence needed to evaluate each criterion outlined in Table 1, the judgmental process required panelists to make a series of ratings (see Appendix B for the specific questions and rating scales):

- Link 1: Extended indicators
  - whether the indicator included academic KSAs
  - how the content of the indicator matches that of the curricular indicator
  - how the performance expectations of the indicator match that of the curricular indicator
  - how the content expectations change over grades

- Link 2: Alternate Assessment Items
  - whether the item measures academic content and to what extended indicator it is aligned
  - how the content of the item matches that of the identified extended indicator
  - how the performance expectation of the item matches that of the identified extended indicator
  - whether the content of the item is age appropriate for the target grade level

Panelists made these ratings first independently and then worked as a panel to come to consensus. These consensus judgments were used as the basis for the subsequent analyses.

Results

The results are summarized in this section with detail included in Appendix C.

Criterion 1 - The content is academic

The first criterion in the LAL evaluation method is to evaluate whether the content of the extended indicators and AA items represent academic (non-foundational) KSAs that are included in the grade level curricular indicators. The results of this analysis (see Table 3) indicate that all extended indicators were rated as academic and most of the AA items were rated as academic (ranging from 90% to 100% across the grade levels). The only exceptions at the item level were one item at grade 4 that was rated as measuring other academic KSAs (not related to the Math curricular indicators) and a few items at grades 7, 8, and 11 that were identified as measuring foundational KSAs. The results at each grade level met (or exceeded) the recommendation from Flowers et. al., (2009) that no more than 10% of the content on an alternate assessment should be foundational.

Table 3. Criteria 1 Results

	Link 1: Extended Indicators			Link 2: AA Items				
	Total	Academic	Percent	Total	Academic	Percent	Academic (Other)	Foundational
3	28	28	100%	25	25	100%	0	0
4	25	25	100%	30	29	97%	1	0
5	20	20	100%	30	30	100%	0	0
6	29	29	100%	30	30	100%	0	0
7	21	21	100%	30	29	97%	0	1
8	23	23	100%	30	27	90%	0	3
11	25	25	100%	30	27	90%	0	3

Based on the results for the first criteria, the questions outlined in Table 1 are addressed below.

Do the extended indicators measure academic content (KSAs)?

- All indicators were identified as measuring academic KSAs.

Do the AA items measure the KSAs that are included in extended indicators?

- Most of the AA items were identified as measuring academic KSAs included in the extended indicators.
- The only exceptions were one item at grades 4 and 7, and three items at grades 8 and 11.

Overall, the results from this first criteria indicate that the extended indicators and AA items are focused on academic KSAs.

Criterion 2 - The content is referenced to the assigned grade level

The second criterion in the LAL evaluation method is to evaluate whether the content of the extended indicators and AA items measures KSAs that are included in the curricular indicators. The extended indicators were created to directly have a 1-to-1 correspondence with the curricular indicators. The results for Link 1 in Table 4 show the distribution of the extended indicators across the four Domains within the Mathematics standards. The AA items were presented to panelists in a "blind" fashion in that the intended indicator alignment was not provided and panelists. Rather, panelists were asked to identify the extended indicator they

felt contained the KSAs required to answer the AA item correctly. As shown in the bottom half of Table 4, the items were aligned to all four Domains with a few identified as not aligning to any extended indicator for that grade.

Table 4. Criterion 2 Results by Domain and Grade Level

	3	4	5	6	7	8	11
Link 1: Extended Indicators to Curricular Indicators							
Indicators	28	25	20	29	21	23	25
1. Number	39%	48%	50%	38%	19%	26%	16%
2. Algebra	18%	12%	15%	28%	52%	39%	36%
3. Geometry	32%	32%	25%	21%	19%	26%	36%
4. Data	11%	8%	10%	14%	10%	9%	12%
Link 2: AA Items to Extended Indicators							
Items	25	30	30	30	30	30	30
1. Number	48%	53%	57%	30%	17%	40%	23%
2. Algebra	12%	17%	17%	37%	50%	20%	27%
3. Geometry	28%	20%	17%	20%	17%	23%	30%
4. Data	12%	7%	10%	13%	13%	3%	13%
Not Aligned	0%	3%	0%	0%	3%	13%	7%

Based on the results for the second criterion, the questions outlined in Table 1 are addressed below.

Do the extended indicators reference KSAs that are included in the associated grade-level curricular indicators?

- Given the direct one-to-one construction of the extended indicators, the panel confirmed the link of all extended indicators to the associated curricular indicators.
- Across all grades, all but four curricular indicators had associated extended indicators.

Do the AA items measure the KSAs that are included in the grade-level extended indicators?

- Most items were aligned to the extended indicators at the targeted grade level. Across all grade levels, the SMEs only identified eight items (4%) that could not be directly aligned to an extended indicator. These eight items are identified in Appendix C for further review by NDE and DRC.

Overall, when creating extended indicators and alternate assessments, it is expected that general curriculum will be narrowed to best accommodate these students. However, the results of this criteria indicate that all four of the mathematics domains were represented in the extended indicators (including 97% of the curricular indicators).

Criterion 3 - The focus of achievement maintains fidelity with the content of the original grade-level standards (content centrality), and when possible, the specified performance (performance centrality).

The third criterion in the LAL evaluation method focuses on to what extent the expectations of the extended indicators and AA items match those in the curricular indicators. Some variation is expected here as the

extended indicators are created to provide a differentiated target for the students taking the NSCAS – AA M. This connection is evaluated based on a match of the content (content centrality) and performance expectations (performance centrality). The results for Links 1 and 2 are shown in Tables 5a and 5b, respectively. For Link 1, most of the extended indicators had a "near" or "far" link to the associated curricular indicator with some exceptions at grades 4, 7, and 11. For Link 2, all AA items has a "near" or "far" link to the content of the aligned extended indicator with the exception of one item at grade 11 (panel indicated they were unsure of the alignment to the identified extended indicator).

Table 5a. Link 1 Indicator Level Content Centrality by Grade and Domain

		3	4	5	6	7	8	11
Number	Near Link	3	0	7	9	1	5	1
	Far Link	8	10	3	1	3	1	3
	No link	0	2	0	0	0	0	0
Algebra	Near Link	0	0	2	8	3	4	1
	Far Link	5	2	1	0	7	5	8
	No link	0	1	0	0	1	0	0
Geometry	Near Link	1	3	5	3	0	3	2
	Far Link	8	4	0	3	3	3	5
	No link	0	1	0	0	1	0	2
Data	Near Link	0	0	0	0	1	1	2
	Far Link	3	2	2	4	1	1	1
	No link	0	0	0	0	0	0	0
Total	Near Link	14%	12%	70%	71%	24%	57%	24%
	Far Link	86%	72%	30%	29%	67%	43%	68%
	No link	0%	16%	0%	0%	10%	0%	8%

Table 5b: Link 2 Item Level Content Centrality by Grade and Domain

		3	4	5	6	7	8	11
Number	Near Link	8	9	17	9	5	7	4
	Far Link	4	7	0	0	0	5	3
	No link	0	0	0	0	0	0	0
Algebra	Near Link	3	4	5	11	14	4	5
	Far Link	0	1	0	0	1	2	2
	No link	0	0	0	0	0	0	1
Geometry	Near Link	7	5	5	6	5	7	7
	Far Link	0	1	0	0	0	0	2
	No link	0	0	0	0	0	0	0
Data	Near Link	3	2	3	4	4	1	3
	Far Link	0	0	0	0	0	0	1
	No link	0	0	0	0	0	0	0
Total	Near Link	84%	69%	100%	100%	97%	73%	68%
	Far Link	16%	31%	0%	0%	3%	27%	29%
	No link	0%	0%	0%	0%	0%	0%	4%

The results for the analysis of performance centrality are shown in Tables 6a and 6b for Link 1 and Link 2, respectively. As noted in the description of this criteria, some performance centrality is expected but likely to a

lesser extent than content centrality as often the creation of extended indicators is done by changing the performance expectation. As shown in Table 6a, the majority of the extended indicators were rated as having "all" or "some" of the performance expectations of the aligned curricular indicator with most of these falling into the "some" category. For Link 2, the results again indicate that the performance expectations for the majority of items matched "all" or "some" of the expectations in the aligned extended indicator with a large percentage being in the "all" category.

Table 6a. Link 1 Indicator Level Performance Centrality by Grade and Domain

		3	4	5	6	7	8	11
Number	All	3	5	7	4	0	2	0
	Some	8	5	2	7	4	4	3
	None	0	2	1	0	0	0	1
Algebra	All	0	1	1	0	1	2	3
	Some	5	1	2	8	7	5	5
	None	0	1	0	0	3	2	1
Geometry	All	1	1	3	1	0	1	0
	Some	8	6	1	5	2	3	5
	None	0	1	1	0	2	2	4
Data	All	0	1	0	0	1	0	1
	Some	2	0	1	4	0	0	2
	None	1	1	1	0	1	2	0
Total	All	14%	32%	55%	17%	10%	22%	16%
	Some	82%	48%	30%	83%	62%	52%	60%
	None	4%	20%	15%	0%	29%	26%	24%

Table 6b. Link 2 Item Level Performance Centrality by Grade and Domain

		3	4	5	6	7	8	11
Number	All	4	8	15	6	4	8	3
	Some	8	8	2	3	1	2	3
	None	0	0	0	0	0	2	1
Algebra	All	3	2	3	9	13	4	5
	Some	0	2	2	2	1	2	2
	None	0	1	0	0	1	0	1
Geometry	All	7	3	2	5	2	7	6
	Some	0	3	3	1	3	0	1
	None	0	0	0	0	0	0	2
Data	All	3	2	3	4	4	1	3
	Some	0	0	0	0	0	0	0
	None	0	0	0	0	0	0	1
Total	All	68%	52%	77%	80%	79%	77%	61%
	Some	32%	45%	23%	20%	17%	15%	21%
	None	0%	3%	0%	0%	3%	8%	18%

Based on the results for the third criterion, the questions outlined in Table 1 are addressed below.

Do the extended indicators maintain fidelity with the content and performance expectations of the original grade-level curricular indicator?

- Across all grade levels, 38% of the extended indicators had a "near link" and 57% had a "far link" to the extended indicators. The remaining extended indicators (5%) had "no link" with the curricular indicator.
- Across all grade levels, 23% of the extended indicators met "all" of the performance expectations of the curricular indicator and 61% met "some" of the performance expectations. The remaining 16% had a "none" rating for performance centrality.

Do the AA items maintain fidelity with the content and performance expectations of the extended indicators?

- Across all grade levels, 85% of the AA items had a near link with the content of the extended indicators, 15% of the AA items had a far link, and 1% had no link to the content of the extended indicator indicating that almost all items were linked to the content in the extended indicator.
- Across all grade levels, 71% of the AA items had an "all" performance centrality rating of "all", 25% had a performance centrality rating of "some" and only 5% had a rating of "none" indicating that most items represented the performance expectations of the associated extended indicator.

Overall, the majority of extended indicators matched the expectations for content and performance from the original curricular indicator. NDE should review those extended indicators identified as "no link" for content centrality or "none" for performance centrality. Similarly, the majority of the AA items had parallel expectations for performance as the aligned extended indicator. NDE should review those AA items identified as "no link" for content centrality or "none" for performance centrality.

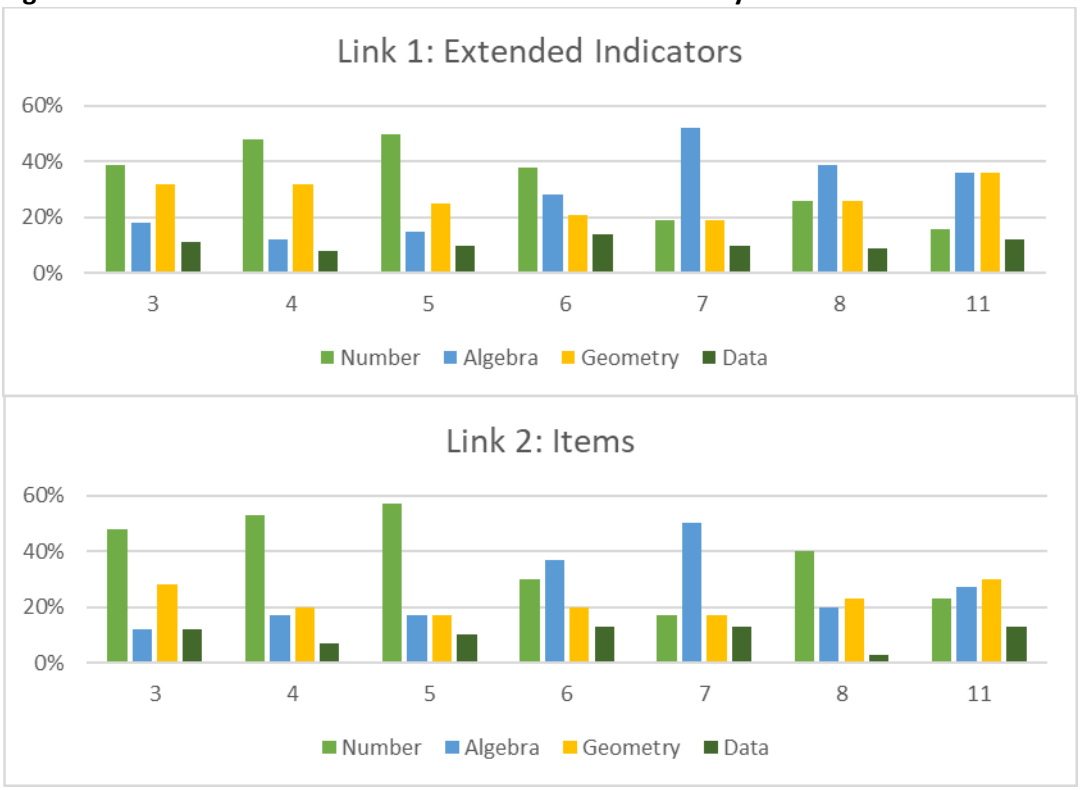
Criterion 4 - The content differs from grade level in range, balance, and DOK but matches high expectations for students with significant cognitive disabilities.

The fourth LAL criterion follows the expectations outlined by Webb (1995; 1997) in his suggested alignment procedures. The first of these is c*ategorical concurrence* or the degree to which the same content categories appear in the extended indicators and the AA items. As shown in Table 7, at each grade level there were items aligned to each of the four mathematics domains. With traditional alignment studies, Webb recommends that each level of reporting have at least six items aligned. However, Flowers et al (2009) note that this expectation may not be realistic given the typical length of alternate assessments. Therefore, the distribution of aligned items was compared to the distribution of extended indicators across the four domains (see Figure 1). This representation highlights the similarities in the distribution of extended indicators and AA items by Domain across grade levels. As examples of this similarity, the large number of extended indicators in the elementary grades are focused on the *Number* domain and the same emphasis is found in the AA items. In the upper grades (middle school and high school), more focus is placed on the *Algebra* and *Geometry* domains which is evident both in the number of extended indicators and AA items.

Table 7. Alignment of Items by Grade and Domain

	Numbers	Algebra	Geometry	Data
3	12	3	7	3
4	16	5	6	2
5	17	5	5	3
6	9	11	6	4
7	5	15	5	4
8	12	6	7	1
11	7	8	9	4

Figure 1. Distribution of Extended Indicators and AA Items by Domain across Grade Level

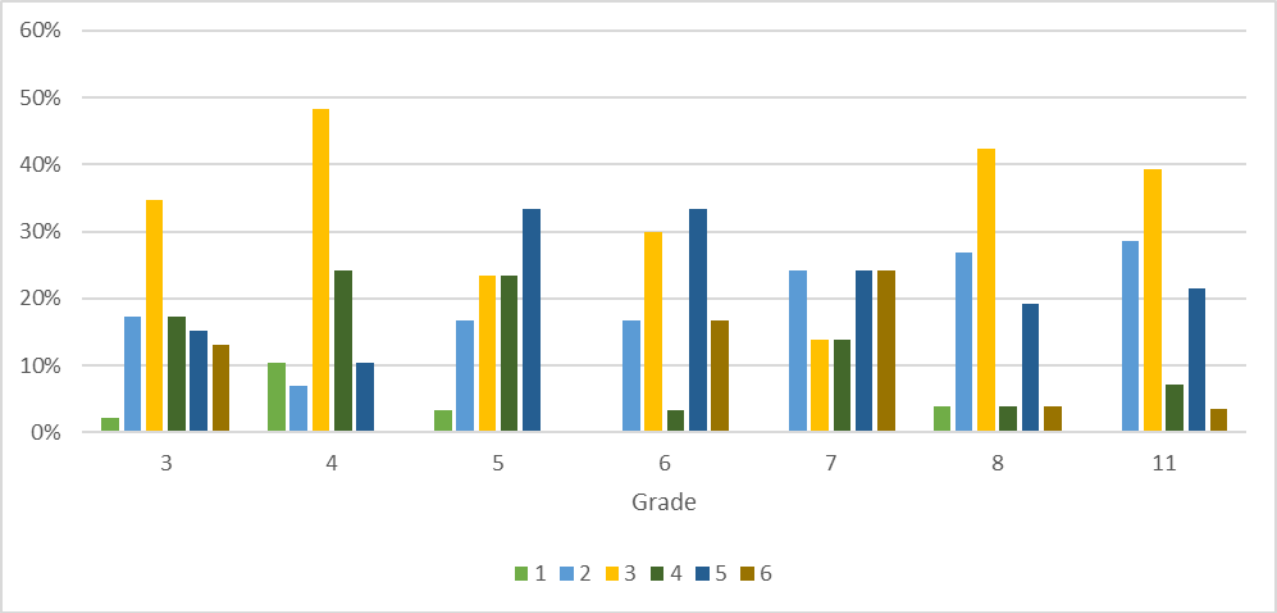


The second expectation in Webb's analysis is an evaluation of depth of knowledge (DOK), or the degree to which the cognitive levels targeted for assessment are required to answer the AA items correctly. Panelists were asked to rate the DOK of each AA item based on Webb's DOK stages for alternate assessment items. In this model, Webb defined six stages of cognitive complexity (1: Analysis through 6: Analysis, Evaluation, see Appendix B). These stages are different than the traditional four levels of DOK purported by Webb but many of the same cognitive processes are in both frameworks. The results of these ratings are shown in Table 8 and graphically in Figure 2. The revised specifications for the NSCAS – AA M include ranges of items by DOK Stage (rather than a specific target stage). However, the expectations could be summarized as an average of one item per extended indicator per DOK Stage 1-4 (i.e., a roughly even distribution across stages 1-4). As shown in Figure 2, the distribution of items by Stage within each grade is somewhat uneven with more items at Stage 3 than other stages. In addition, items representing Stages 5 and 6 were identified in the assessments although these stages are beyond the expectations outlined in the NSCAS – AA M specifications. Webb suggests that a goal is for 50% of the items to be at or above the targeted DOK as indicated in the specifications. Given that 70% of the items were at or above a DOK Stage 3, this expectation can be considered met.

Table 8. Number of items at each DOK Stage by Grade

	1	2	3	4	5	6
3	1	8	16	8	7	6
4	3	2	14	7	3	0
5	1	5	7	7	10	0
6	0	5	9	1	10	5
7	0	7	4	4	7	7
8	1	7	11	1	5	1
11	0	8	11	2	6	1

Figure 2. Distribution of item DOK by Grade and Stage.



The last two expectations from Webb are focused on the distribution of content across the full set of expectations. Given the length of the NSCAS – AA M exam, the focus of this analysis was on the alignment to the GLEs within a Domain. The *range of knowledge* expectation evaluates how many GLEs (within each Domain) are represented on the test form to ensure the different areas of the Domain are being measured. At a deeper level, the *balance of representation* expectation evaluates how the content aligned to a Domain is distributed across the subsumed GLEs. The results for these two criteria are shown in Table 9. For each grade and domain, the range of knowledge results indicate what percent of GLEs within that domain have at least one item aligned. From these values, the next column indicates whether the results meet the range of knowledge expectation (50% or more GLEs have an aligned item).

For each grade and domain, the balance of representation value is calculated using the following formula:

$$BOR\ Index = 1 - \frac{\sum_{K=1}\left(abs\left[\frac{1}{Number\ of\ GLEs\ (in\ domain)\ with\ items\ aligned} - \frac{Number\ of\ item\ hits\ to\ GLE\ (k)}{Number\ of\ item\ hits\ to\ the\ Domain} \right] \right)}{2}$$

The results are then evaluated against the expectation recommended by Webb (values should be greater than or equal to 0.70). Although both the range of knowledge and balance of representation expectations are included in this analysis, the recommendations from the literature are mixed as to whether these are reasonable goals for alternate assessments (USDOE, 2015).

Table 9. Results of Range of Knowledge and Balance of Representation Results by Grade and Domain

		Range of Knowledge		Balance of Representation	
Grade	Domain	% of GLEs with 1 or more aligned items	Met Criterion?	BOR	Met Criterion?
3	Number	100%	Yes	0.67	No
3	Algebra	67%	Yes	0.67	No
3	Geometry	100%	Yes	0.93	Yes
3	Data	50%	Yes	0.50	No
4	Algebra	100%	Yes	0.69	No
4	Geometry	100%	Yes	0.90	Yes
4	Data	100%	Yes	0.67	No
4	Algebra	100%	Yes	1.00	Yes
5	Algebra	100%	Yes	0.91	Yes
5	Geometry	67%	Yes	0.67	No
5	Data	100%	Yes	0.73	Yes
5	Algebra	100%	Yes	1.00	Yes
6	Algebra	100%	Yes	0.94	Yes
6	Geometry	100%	Yes	0.77	Yes
6	Data	100%	Yes	1.00	Yes
6	Algebra	100%	Yes	1.00	Yes
7	Algebra	100%	Yes	1.00	Yes
7	Geometry	100%	Yes	0.93	Yes
7	Data	100%	Yes	0.70	Yes
7	Algebra	100%	Yes	0.75	Yes
8	Algebra	100%	Yes	0.83	Yes
8	Geometry	100%	Yes	0.83	Yes
8	Data	100%	Yes	0.76	Yes
8	Algebra	50%	Yes	0.50	No
11	Algebra	100%	Yes	0.64	No
11	Geometry	100%	Yes	0.88	Yes
11	Data	100%	Yes	0.89	Yes
11	Algebra	50%	Yes	0.50	No

Based on the results for the fourth criteria, the questions outlined in Table 1 are addressed below.

Categorical concurrence: Do the same categories of content appear in the extended indicators and the AA items?

- The results show that all four mathematics domains were represented on each of the grade level assessments and the distribution of extended indicators across the four domains paralleled the distribution of aligned items.

Depth of knowledge: Do the AA items target the same range of cognitive levels as the extended indicators?

- The NSCAS – AA M tables of specifications call for a distribution of items across DOK stages 1-4. The AA items were coded to show a distribution across stages but with more emphasis at stage 3 and representation of higher stages including 5 and 6 which were not part of the original specifications.
- Although there is some difference between the targeted distributions and the actual distributions, Webb suggests that items should be at/above the targeted level. The evidence from this study suggests each grade level exam meets this expectation.

Range of knowledge: Do the AA items target the same range of knowledge as the Domains and GLEs?

- Overall, all Domains (across all grade levels) met the range of knowledge expectation and, on average, 92% of the GLEs had at least one item aligned indicating the content was aligned to a range of expectations.

Balance of representation: Are the AA items distributed across the GLEs within each domain?

- Across all grade levels, most (68%) of the domains met the expectations for balance of representation (the content aligned to a domain was distributed across the GLEs). NDE should review those domains within each grade level that did not meet the expectation noting that this expectation within Webb's framework should not be overinterpreted. If there is a curricular, instructional, or assessment design rationale for why the representation would not be met, that could support maintaining the current sampling of each domain.

Criterion 5 - There is some differentiation in content across grade levels.

The final LAL criterion is an evaluation as to how the content expectations (indicators) change across grade level and then the age appropriateness of the AA items. For each indicator, panelists were asked to determine how it compared to the similar lower grade indicator by determining if it represented:

- a broader classification of the target skill or knowledge
- a deeper mastery of the target skill or knowledge
- a skill that for which the prerequisite was included at a lower grade standard
- a new skill or knowledge that was not included in the lower grade levels
- a skill or knowledge that is identical to one that was included in a lower grade level

From an assessment development perspective, the first four of the above classifications are considered acceptable as they indicate higher expectations for higher grade levels whereas "identical" indicate a lack of escalation of expectations. The results of this analysis (see Table 10) indicate that the majority of indicators represented a *broader* expectation of the target skill or knowledge or a *new* skill or knowledge. Overall, all but two indicators suggested a change between grades.

Table 10. Link 1 - Differentiation across grades by classification.

Grade	Broader	Deeper	Prerequisite	New	Identical
3	--	--	--	--	--
4	9	6	1	8	1
5	10	1	0	9	0
6	8	6	0	15	0

Grade	Broader	Deeper	Prerequisite	New	Identical
7	11	1	2	7	0
8	6	3	0	14	0
11	3	0	2	19	1

At the item level, the panelists were asked to evaluate each item to determine if it was age appropriate. Specifically, each item was classified as either adapted from age specific content, age neutral, or inappropriate for the target age group. The results of this analysis (Table 11) indicate that most items were age neutral (appropriate for all students) or from age specific content. Only two items at each of grades 8 and 11 were identified as inappropriate for the target age group. NDE should review these items to determine possible revisions.

Table 11. Link 2 – Age appropriateness of items

	Age specific	Neutral	Inappropriate
3	5	20	0
4	9	20	0
5	0	30	0
6	0	30	0
7	0	29	0
8	12	12	2
11	13	13	2

Based on the results for the fifth criterion, the questions outlined in Table 1 are addressed below.

How does the content specified in the extended indicators change across grade levels?

- Most of the extended indicators represented a broader application of a target knowledge or skill or introduction of a new application or skill.
- There were only two extended indicators rated as identical (one at grade 4, one at grade 11) that should be reviewed by NDE.

Are the AA items age appropriate for each grade level?

- Almost all AA items were rated as age neutral or age specific.
- The four items identified as inappropriate for the targeted age group (2 at grade 8, 2 at grade 11) should be reviewed by NDE.

Process Evaluation

The results from the process evaluation are shown in Table 12. Overall, the panelists felt the training was successful, there was an appropriate amount of time for the training, and they were confident in their ratings of the extended indicators and items. The only consistent comment was that the third panel would have appreciated only completing the alignment process for two grade levels (versus three).

Table 12. Process Evaluation Results

Evaluation Question	3 & 4	5 & 6	7, 8, & 11
How successful was the training session in preparing you for making the alignment judgments?	4	4	3

Evaluation Question	3 & 4	5 & 6	7, 8, & 11
4 = Totally Successful to 1 = Totally Unsuccessful			
How appropriate was the amount of time dedicated to each training component? *3 = Too much time to 1 = Not enough time*	2	2	2
How confident are you in your judgments of the Extended Indicators? *3 = Very confident to 1 = Not very confident*	3	3	3
How confident are you in your judgments of the items? *3 = Very confident to 1 = Not very confident*	3	3	3
Comments - It was a great experience! - good experience for me - Once standards are fully released, this activity would be great for teachers! Thanks! - Even though we had the wrong items we were able to get it done - This was a great experience! Thanks! - Thank you for this opportunity! - Disappointing that we wanted most of the morning on Thursday. Try to think about equity of work - adding another group just to do 11th grade would have been helpful as well as some general education content experts. - I learn by doing but Susan was helpful as we went along. - [Regarding the amount time] - it was the right amount of time considering we had 3 levels to go through. - Give one group 11th grade and another 7 and 8 - I loved getting to participate! I wasn't even aware of the "extended indicators" and feel that it will be greatly beneficial to my teaching! Thank you for the opportunity. - Have a separate group for 11th grade as it is more in-depth - Show some examples - Maybe another group just for 11th grade			

Evaluating the Alignment Study

To evaluate the alignment study, we applied the framework suggested by Davis-Becker and Buckendahl (2013). Within this framework, the authors suggested four sources of evidence that should be considered in the validation process: procedural, internal, external, and utility. Threats to validity that were observed in these areas should mitigate policymakers' judgments regarding the usefulness of the results and the validity of the interpretation. Evidence within each of these areas that was observed in this study is discussed here.

Procedural

Procedural evidence focuses on panelist selection and qualifications, choice of methodology, application of the methodology, and panelists' perspectives about the implementation of the methodology. For this study, the panel that was recruited included experienced educators in various roles from across the state. In addition, the

panelists were independent of any development and validation activities for the NSCAS – AA M. The LAL methodology is one that was designed specifically for alternate assessments.

Internal

The internal evidence for alignment studies can be evaluated by examining the consistency of panelists' ratings and the convergence of the recommendations. For this study, the rating tasks and decision rules were based on consensus judgments which requires greater agreement among panelists. Although the results should not be interpreted as unanimous support by the panelists, the panelists worked well together in evaluating differences of opinion to determine the most appropriate consensus judgment.

External

External validity evidence comes a separate source whereby the alignment results are confirmed through evidence collected outside the alignment study. For this study all data were collected during the workshop. If external evidence is desired, NDE could conduct a broader survey of educators to react to the results.

Utility

Evidence of utility is based on the extent to which the summative and formative feedback can be used to inform policy and operational decisions related to the interpretation of scores. We believe that the summative information from the study provides the evidence necessary for Nebraska to move forward with the use of the new NSCAS – AA M as their statewide alternate assessment for Mathematics. There are a few specific findings, identified through the report, that we recommend NDE review for future test development activities.

References

Davis-Becker S. & Buckendahl, C. W. (2013). A proposed framework for evaluating alignment studies. *Educational measurement: Issues and practice, 32*(1), 23-33.

Flowers, C., Wakeman, S., Browder, D., & Karvonen, M. (2009). Links for academic learning (LAL): A conceptual model for investigating alignment of alternate assessments based on alternate achievement standards. *Educational measurement: Issues and practice*, 28(1), 25-37.

U.S. Department of Education, Office of Special Education Programs (2016). *Validity Evidence*. Accessed from https://osepideasthatwork.org/node/9

Webb, N. L. (1995, April). *Issues Related to Judging the Alignment of Curriculum Standards and Assessments*. Paper presented at the annual meeting of the American Educational Research Association. Montreal, CA.

Webb, N. L. (1997). Criteria for alignment of expectations and assessments in mathematics and science education (Council of Chief State School Officers and National Institute for Science Education Research Monograph No. 6). Madison, WI: University of Wisconsin, Wisconsin Center for Education Research.

Appendix A: Study Materials

The documents below are the materials used in the judgmental process throughout the study.

Study Resources

Training slides.pdf

DOK Resource.pdf

Nebraska Mathematics Standards with Extended Indicators

Math_Extended_Indicators_2017_Final.pdf

NSCAS – AA M Tables of Specification

G11-NSCAS-Alternate-Math-TOS.pdf

G8-NSCAS-Alternate-Math-TOS.pdf

G7-NSCAS-Alternate-Math-TOS.pdf

G6-NSCAS-Alternate-Math-TOS.pdf

G5-NSCAS-Alternate-Math-TOS.pdf

G4-NSCAS-Alternate-Math-TOS.pdf

G3-NSCAS-Alternate-Math-TOS.pdf

Appendix B: Rating Scales

The following tables provide the rating scales used by panelists to make their ratings for links 1 and 2.

Academic (Criterion 1)

Code	Link 1: Extended Indicators	Link 2: Items
Yes	The extended indicator includes KSAs that are included in the grade-level Mathematics standards/indicators	The KSAs required to answer the item correctly can be found in the grade-level Mathematics extended indicators [identify indicator]
No – F	The extended indicator describes foundational KSAs (	The KSAs required to answer the item correctly are foundational.
No – 0	The extended indicator includes other non-academic KSAs (e.g., following directions)	The KSAs required to answer the item are non-academic and not foundational.
No – A	The extended indicator includes academic KSAs from a different content area (e.g., Reading)	The KSAs required to answer the item are academic but aligned to a different content area (e.g., Reading)

Content Centrality (Criterion 4)

Code	Link 1: Extended Indicators	Link 2: Items
Near Link	Most (50% or more) of the content in the general grade level indicator is present in the extended indicator.	The item content is contained within the identified indicator.
Far Link	Some (between 1% and 49%) of the content in the general grade level indicator is present in the extended indicator.	Some of the item content is included in the identified indicator
No link	None (0%) of the content in the general grade level indicator is present in the extended indicator.	The item content is not really included in the indicators

Performance Centrality (Criterion 4)

Code	Link 1: Extended Indicators	Link 2: Items
All	The level of expected performance for the extended indicator matches all of the performance required in the grade level content indicator	The performance of the item/task is identical to the performance of the content standard
Some	The level of expected performance for the extended indicator matches some of the performance required in the grade level indicator	The performance of the item/task partially matches the performance of the content standard (may occur when two different performances are asked in the content standard).
None	The level of expected performance for the extended indicator does not match any of the performance required in the grade level indicator	The performance of the item/task does not match the performance of the content standard

Depth of Knowledge (Criterion 4)

Level	Link 2: Items
1	Attention (touch, look, vocalize, respond, attend)
2	Memorize/recall (list, describe (facts), identify, state, define, label, recognize, record, match, recall, relate)
3	Performance (perform, demonstrate, follow, count, locate, read)
4	Comprehension (explain, conclude, group/categorize, restate, review, translate, describe (concepts), paraphrase, infer, summarize, illustrate)
5	Application (compute, organize, collect, apply, classify, construct, solve, use, order, develop, generate, interact with text, implement)
6	Analysis, Synthesis, Evaluation (pattern, analyze, compare, contrast, compose, predict, extend, plan, judge, evaluate, interpret, cause/effect, investigate, examine, distinguish, differentiate, generate)

Change Across Grades (Criterion 5)

Level	Link 1: Indicators
Broader	the higher-grade indicator reflects broader application of the target skill or knowledge
Deeper	the higher-grade indicator reflects deeper mastery of the target skill or knowledge
Prerequisite	the lower-grade indicator reflects a different prerequisite skill for mastery of the higher-grade indicator
New	the higher-grade indicator is a new skill or knowledge covered at prior grades
Identical	the higher-grade indicator appears identical to one of the lower grade indicators

Age Appropriate (Criterion 5)

Level	Link 2: Items
1	Adapted from age-specific content
2	Age neutral
3	Inappropriate for age group

Appendix C: Detailed Results

The file below contains the consensus rating for all extended indicators and NSCAS – AA M items.

Item and Indicator Results.xlsx

As noted in the report, there are a few areas where NDE should review the findings of the alignment panel and determine if any revisions are needed to the extended indicators or NSCAS – AA M items. These are identified below by criteria.

Criteria 1 and 2: The table below lists all AA items that were identified as either representing non-academic KSAs or not having directly alignment to one of the grade-level extended indicators.

Grade	Item	Academic	Alignment	Notes
4	8	No	None	This question doesn't match any extended indicator.
7	16	No	None	
8	3	No	None	8.2.2.b requires two step inequality; this is a statement
8	6	Yes	None	Can't find a clear indicator to match
8	9	No	None	The only thing this question is asking is if the student knows what "longest" means. Can't find an indicator to relate to.
8	18	No	None	8.2.2.b requires two step inequality; this is a statement; also a repeat of #3
11	3	No	11.2.2.d	
11	8	No	None	
11	13	No	None	

Criterion 3: The table below lists all extended indicators that were rated as having either “no link” for content centrality or “none” for performance centrality.

Domains	Indicator	Content Centrality	Performance Centrality
4	3.4.1.a	Far Link	None
1	4.1.1.c	No link	None
1	4.1.2.e	No link	None
2	4.2.1.a	No link	None
3	4.3.1.d	No link	None
4	4.4.1.a	Far Link	None
1	5.1.1.d	Far link	None
3	5.3.1.c	Near link	None
4	5.4.2.b	Far link	None
2	7.2.1.a	No link	None
2	7.2.3.e	Far Link	None

Domains	Indicator	Content Centrality	Performance Centrality
2	7.2.3.f	Far Link	None
3	7.3.1.a	Far Link	None
3	7.3.3.c	No link	None
4	7.4.3.c	Far Link	None
2	8.2.1.c	Far Link	None
2	8.2.3.b	Far Link	None
3	8.3.2.a	Far Link	None
3	8.3.3.d	Far Link	None
4	8.4.1.a	Far Link	None
4	8.4.2.a	Near Link	None
1	11.1.1.a	Far Link	None

Domains	Indicator	Content Centrality	Performance Centrality
2	11.2.2.g	Far Link	None
3	11.3.1.c	Far Link	None
3	11.3.1.d	No link	None

Domains	Indicator	Content Centrality	Performance Centrality
3	11.3.2.c	Far Link	None
3	11.3.2.d	No link	None

The table below provides a list of all items that were identified as either having "no link" to the identified extended indicator or "none" of the performance expectations in the indicator.

Grade	Item	Aligned Indicator	Content Centrality	Performance Centrality
4	26	4.2.3.a	Far Link	None
7	21	7.2.2.c	Far link	None
8	2	8.1.1.a	Far Link	None
8	14	8.1.1.a	Far Link	None
11	10	11.1.2.b	Far Link	None
11	19	11.3.2.e	Far Link	None
11	20	11.4.3.c	Far Link	None
11	26	11.3.3.e	Far Link	None
11	29	11.2.1.c	No link	None

Criterion 4: The table below provides a list of the domains that did not meet the balance of representation criteria.

Grade	Domain	BOR	Meet BOR Criteria
3	3.1	0.67	No
3	3.2	0.67	No
3	3.4	0.50	No
4	4.1	0.69	No
4	4.3	0.67	No
5	5.2	0.67	No
8	8.4	0.50	No
11	11.1	0.64	No
11	11.4	0.50	No

Criterion 5: The table below lists the indicators identified as "identical" to the previous grade.

Grade	Domains	Change over Grades
4	1	Identical
11	1	Identical

The table below lists the items identified as inappropriate for the targeted age group.

Grade	Item	Academic	Aligned Indicator	Content Centrality	Performance Centrality	DOK	Age Appropriate
8	10	Yes	8.1.2.c	Near Link	All	3	3 - Inappropriate
8	17	Yes	8.2.1.b	Far Link	Some	3	3 - Inappropriate
11	2	Yes	11.2.2.g	Far Link	Some	3	3 - Inappropriate
11	30	Yes	11.4.3.b	Near Link	All	3	3 - Inappropriate